Rolling Deck to Repository (R2R):

Collaborative Development of Linked Open Data

R2R Program

Alpha Helix (retired)	Atlantic Explorer	Atlantis	Blue Heron
Cape Hatteras (retired)	Clifford A. Barnes	Corwith Cramer	Endeavor
F.G. Walton Smith	Healy	Hugh R. Sharp	Ka'imikai-o-Kanaloa
Kilo Moana	Knorr	Laurence M. Gould	Marcus G. Langseth
Maurice Ewing (retired)	Melville	Moana Wave (retired)	Nathaniel B. Palmer
New Horizon	Oceanus	Pelican	Point Sur
Polar Sea	Robert C. Seamans	Robert Gordon Sproul	Roger Revelle
Savannah	Sikuliaq	Thomas G. Thompson	Wecoma (retired)

Mission:
Provide uniform stewardship of routinely-collected environmental sensor data from the US academic research fleet

26 active service vessels

~400 primary sensor systems

~400 cruises/year

Core Services

- Inventory cruises in a central catalog

- Upload sensor data from vessels; organize and document; deposit to long-term (NOAA) archives

- Assess quality of selected datasets

- Create quality-controlled products incl. shiptrack navigation, real-time met/ocean, geophysical profiles

- Provide tools for at-sea event logging

R2R Catalog

As of Sep. 2013
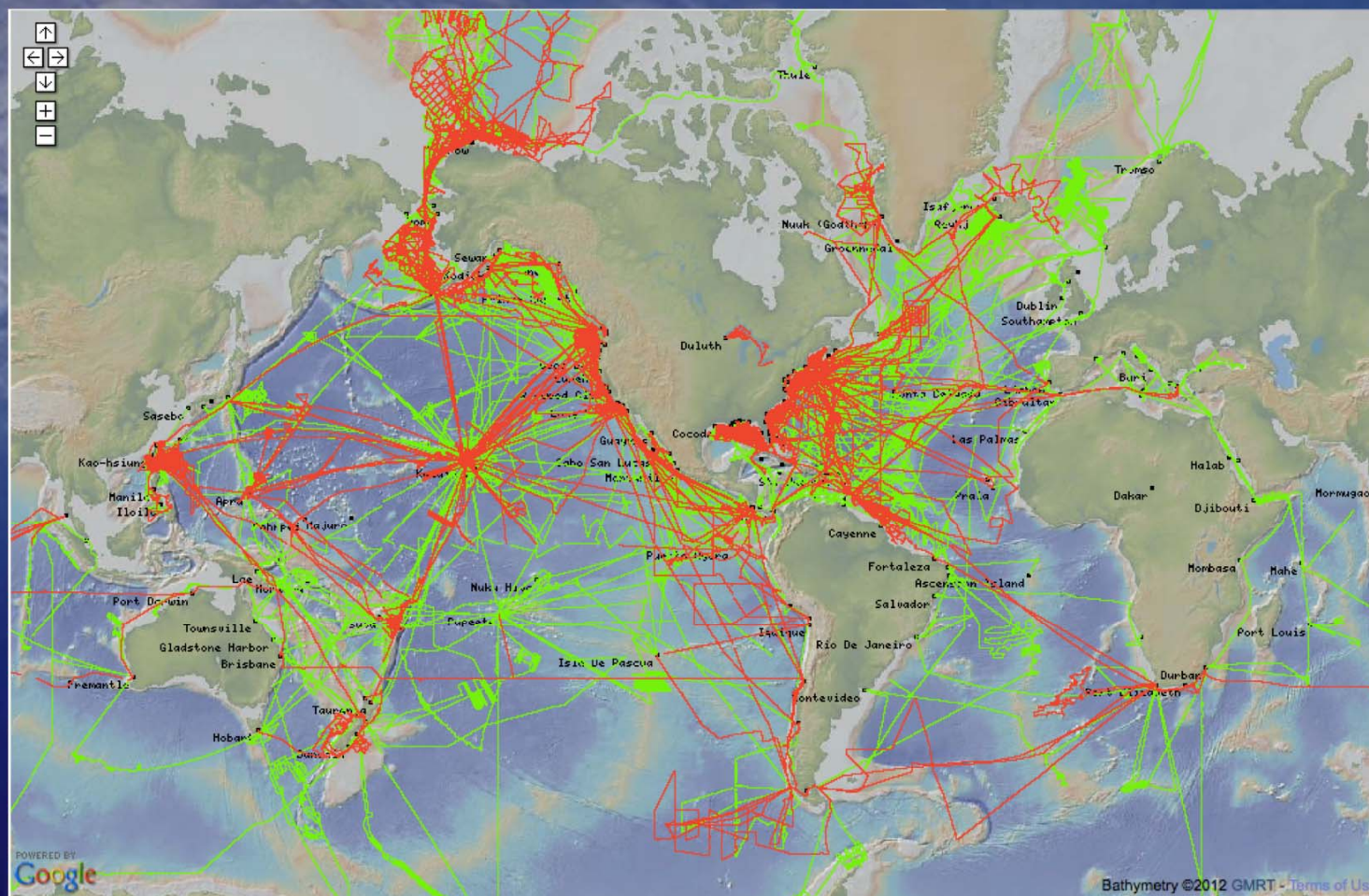
— 1,772 New cruises since R2R launch in 2009

— 1,479 Additional / pre-2009 cruises received from vessel operators
    or other databases (LDEO, SIO, WHOI)

3,251 Total cruises

3

Open Access

Goal: Provide R2R Catalog content that is

- Authoritative
  (complete and correct, with persistent/unique identifiers)

- Accessible
  (lightweight, standard interfaces)

- Interoperable
  (mapped to international/community vocabularies)

Open Access *(cont.)*

Publish Catalog content in multiple formats, per user demand:

1. ISO 19115-2 XML Metadata records

2. OGC Web Services

3. W3C "Linked Data" + SPARQL endpoint

all using standard vocabularies

Linked Open Data?

Publish content using World Wide Web Consortium (W3C) standards and community best practices:

- Each resource (Cruise, Vessel, Instrument, Dataset, File, Sample, Person, Place, etc) has a unique identifier (URI)

- Each URI returns content via HTTP in Resource Description Framework (RDF)

- Resources can be searched using RDF Query Language (SPARQL)

- Resources contain links (URIs) to *additional/related content* in other systems

- Content is published under an open license eg. CC 3.0

R2R LOD Implementation

Deployed D2R Server (*http://d2rq.org/*) open-source package that maps existing SQL database to RDF on the fly.

Provides

- (backend) SPARQL query endpoint for machines

- (frontend) Pubby browser for humans

So far, performance is good for ~tens of thousands of triples.

http://linked.rvdata.us/

R2R LOD Implementation *(cont.)*

Top-level browse over all available resources in the Catalog

List of Datasets (collections of data files from sensor/ system X on vessel Y during cruise Z)

One dataset, with file manifest and download link

http://linked.rvdata.us/

LOD Partners

- Integrated Ocean Drilling Program (IODP)

- Rolling Deck to Repository (R2R) Program

- Biological & Chemical Oceanography Data Management Office (BCO-DMO)

- British Oceanographic Data Centre (BODC)

- Index to Marine and Lacustrine Geological Samples (IMLGS) ★
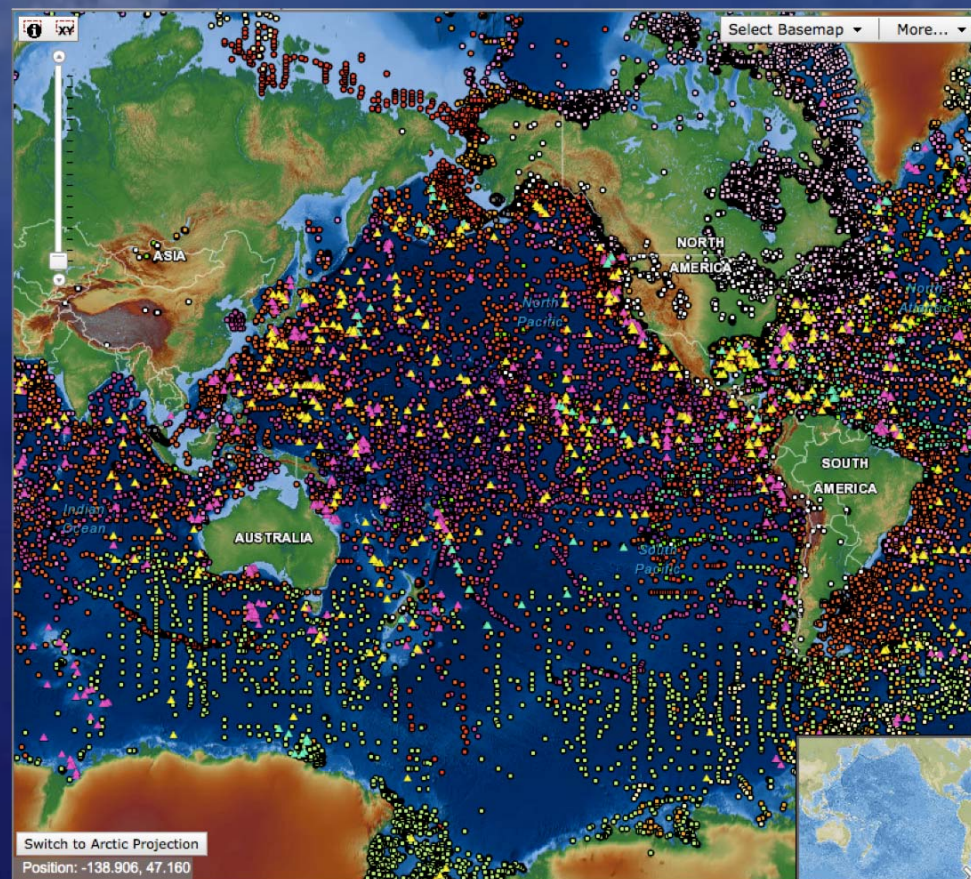
- InterRidge Hydrothermal Vents Database ★

★ New in 2013

LOD Partners *(cont.)*

Index to Marine and Lacustrine Geological Samples (IMLGS)

- Global database of rock and sediment cores, dredges, grabs from seafloor/lakebeds

- Links samples to:
  - ICES Platform Codes
  - R2R Cruise IDs
  - Intl Geo Sample Numbers (IGSNs)

  where known

R2R <> IMLGS

Cruise at IMLGS

Cruise at R2R

http://linked.seabedsamples.org/

*Intl Geo Sample
Number
(link to registry)*

*Platform
(link to ICES
ship code)*

LOD Vocabularies

http://vocab.nerc.ac.uk/

Work ongoing to replace R2R vocabularies with terms from NERC Vocabulary Server (NVS) eg.

- Map vessels to SeaVoX platform categories (L06) and ICES platform codes (C17)

- Map ports to SeaDataNet keyword types (L19) and gazetteer (C38)

- Map devices to SeaDataNet device categories (L05) and BODC device catalog (L22)

Future work: parameters (P02), units (P06), activities (C77)

LOD Geometries

Hybrid approach to publish geospatial information:

- Encode spatial extent of Cruise track (bounding box) using *wgs84_pos:long* and *wgs84_pos:lat*

- Encode control points for Cruise track (linestring) using *geo:asWKT* from GeoSPARQL vocabulary
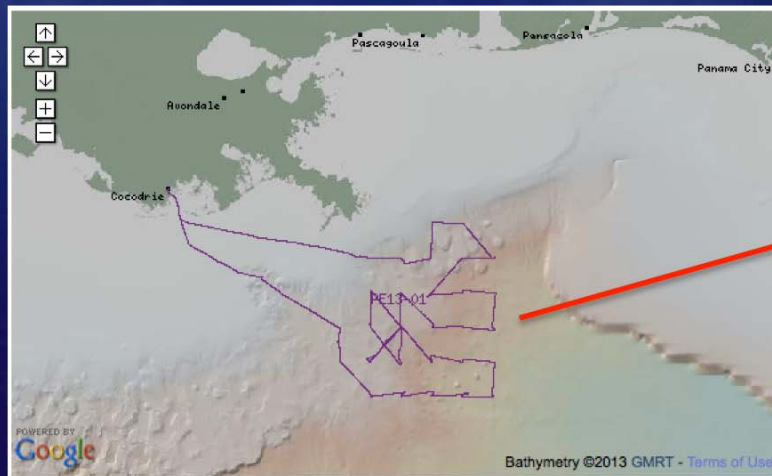
Example: R/V Pelican cruise PE13-01 track

14

ISO Metadata

- R2R publishes *fieldSession* (Cruise) level records; eventually *series* (Dataset) and *dataset* (File/Sample) as well

- Embed shiptrack geometry as GML object in Cruise record

- Replace *gco:CharacterString* (text) with *gmx:Anchor* (LOD URI) for controlled vocabulary terms where possible eg.

```
<gmd:MD_Identifier>
    <gmd:code>
        <gco:CharacterString>KM0715</gco:CharacterString>
    </gmd:code>
</gmd:MD_Identifier>

<gmd:MD_Identifier>
    <gmd:code>
        <gmx:Anchor xlink:href="http://linked.rvdata.us/resource/cruise/KM0715"
xlink:actuate="onRequest">KM0715</gmx:Anchor>
    </gmd:code>
</gmd:MD_Identifier>
```

OGC Services

Publish Cruise track control points as Web Feature Service (WFS) with embedded LOD URIs

http://get.rvdata.us/wfs?service=wfs&version=1.0.0&request=getcapabilities
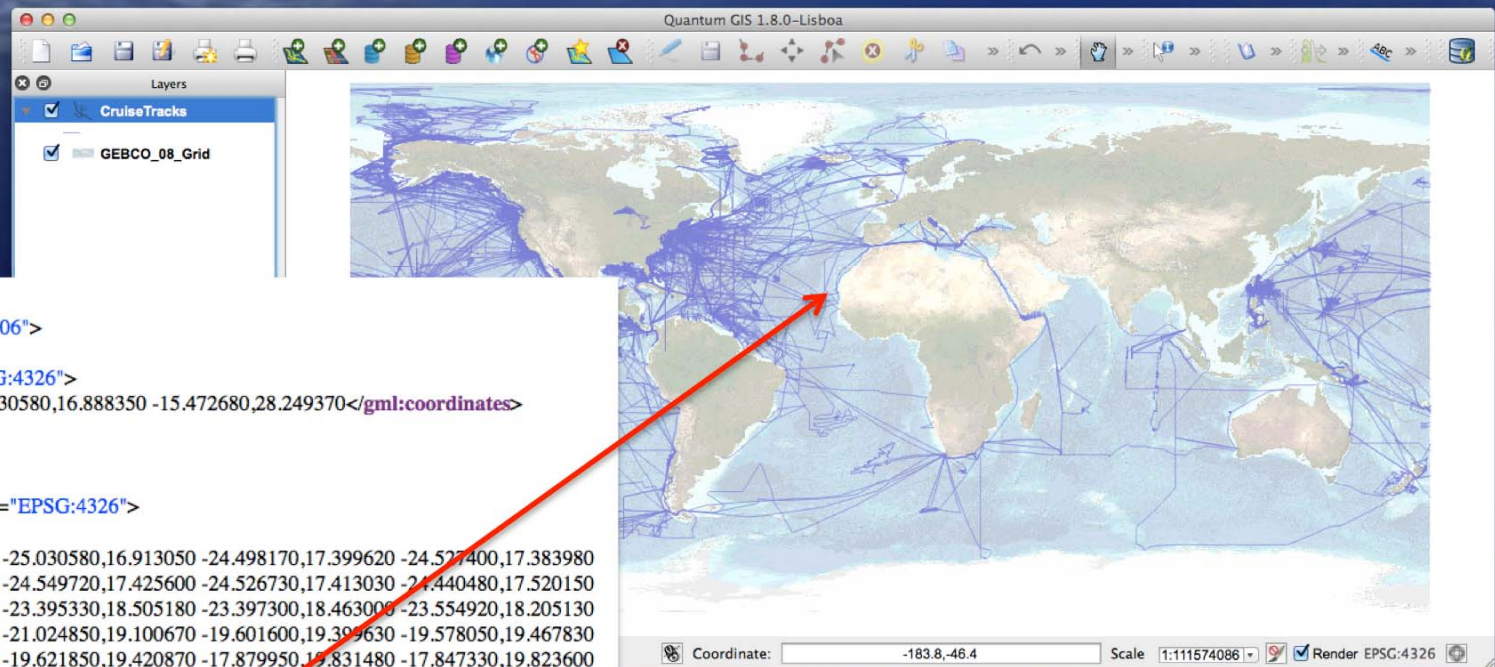

```
- <gml:featureMember>
  - <ms:CruiseTracks fid="41770506">
    - <gml:boundedBy>
      - <gml:Box srsName="EPSG:4326">
          <gml:coordinates>-25.030580,16.888350 -15.472680,28.249370</gml:coordinates>
        </gml:Box>
      </gml:boundedBy>
    - <ms:msGeometry>
      - <gml:LineString srsName="EPSG:4326">
        - <gml:coordinates>
            -25.005000,16.888350 -25.030580,16.913050 -24.498170,17.399620 -24.527400,17.383980
            -24.500900,17.400530 -24.549720,17.425600 -24.526730,17.413030 -24.440480,17.520150
            -23.912350,17.992720 -23.395330,18.505180 -23.397300,18.463000 -23.554920,18.205130
            -21.484170,19.018730 -21.024850,19.100670 -19.601600,19.399630 -19.578050,19.467830
            -19.596120,19.386820 -19.621850,19.420870 -17.879950,19.831480 -17.847330,19.823600
            -17.701220,19.888130 -17.449880,19.933680 -17.508020,19.918370 -17.544950,19.945230
            -15.586400,28.249370 -15.472680,28.249120
          </gml:coordinates>
        </gml:LineString>
      </ms:msGeometry>
      <ms:CruiseID>OC449-03</ms:CruiseID>
      <ms:CruiseURI>http://linked.rvdata.us/resource/cruise/OC449-03</ms:CruiseURI>
    </ms:CruiseTracks>
  </gml:featureMember>
```
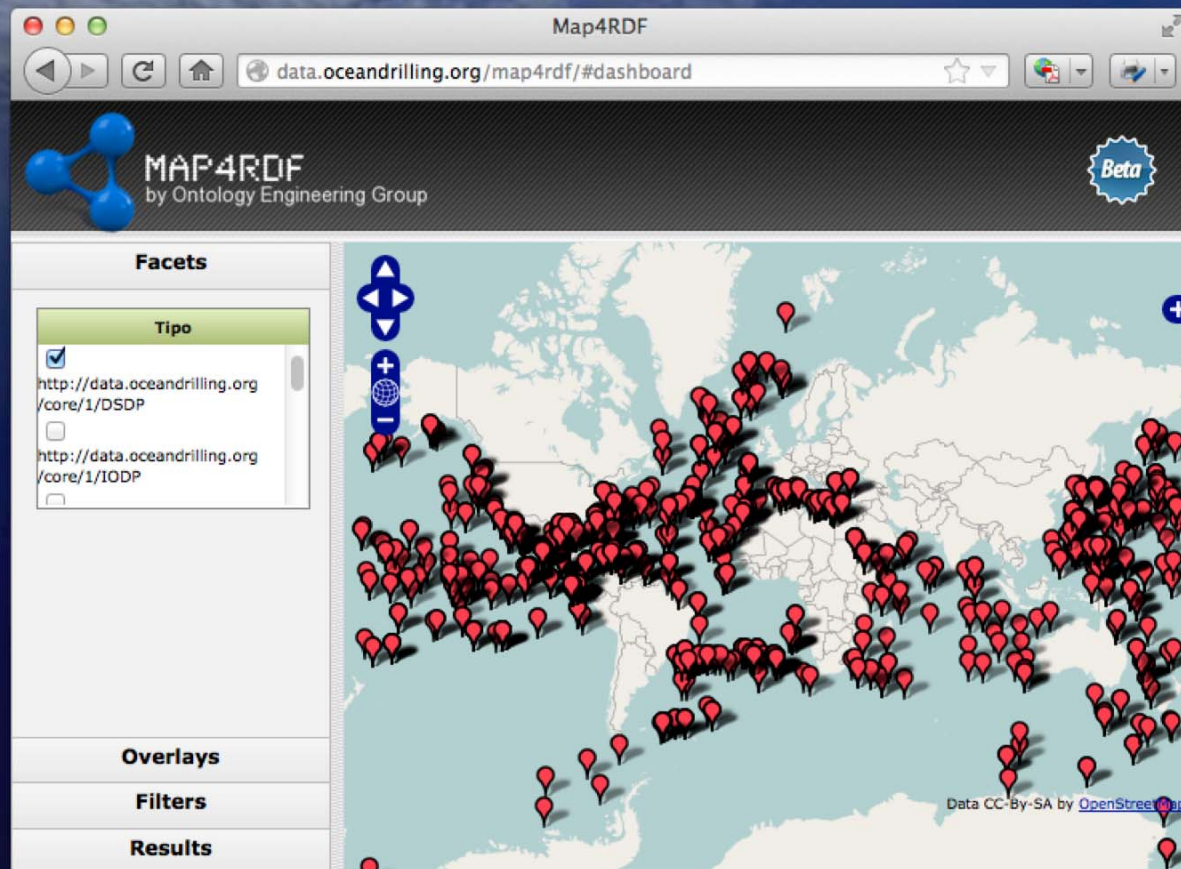
Cruise OC449-03 trackline geometry

R2R Catalog displayed in QGIS desktop client

16

Next Steps

- Deploy Web map-based browser over LOD content/partners

http://data.oceandrilling.org/map4rdf/

Next Steps *(cont.)*

- Implement new W3C Data Set Catalog (DCAT) vocabulary eg.

    dcat:Catalog , dcat:Dataset , dcat:Distribution

- Extend /improve query capabilities

    - Upgrade to new GeoSPARQL standard ("get tracks/ samples within radius X of point Y")

    - Deploy OpenSearch - publish a set of standard queries +responses

- Participate in Ocean Data Interoperability Platform (ODIP) int'l workshops and prototypes

info@rvdata.us